

Data Mining and Analysis for Managers
MGMT 635101
Fall 2021
CLASS GROUP PROJECT

TEAM PLATINUM:

Ayush Vashistha
Prajwal Pandarinath Embar
Harish Naricchetty
Yeshwanth Sai Jayavarapu

Prepared for:
Professor Wayne G Fox

1. Input-data Analysis

Based on the input file given, we must modify the data so that it can be used for data mining. We must look for errors and mistakes and clean the data and remove unimportant variables.

1. In the first instance, the workbook focuses on the **SouthEast** region of the **Region** variable.

Region	Store ID	Population	Store Opening	Sales		Product Purchase	Coupon Received	Avg Monthly			Store Location	Staff Age	Sales BackGround	Loyalty Card	% Sales Staff College	Parking Places	Total Sales	
				Staff/1000	Sq ft			Facebook	Ship to Store	Ship to Store								
SouthEast	1013	62,684	7/13/2012	20	8,202	4 Yes	WorkOut	PC	1%	44,506	Ship	Mall	22	Marketing	Yes	100%	24	\$207,511
SouthEast	1014	96,631	7/10/2010	14	5,628	1 No	WorkOut	PC	50%	47,927	Ship	Stand Alone	27	Marketing	No	105%	NA	\$142,388
SouthEast	1015	62,484	1/26/2009	18	3,100	3 No	Biking	PC	18%	47,248	Ship	Mall	28	Marketing	Yes	84%	12	\$78,430
SouthEast	1016	40,108	12/27/2012	14	3,729	1 Yes	Biking	PC	16%	12	Ship	Stand Alone	25	Operations	Yes	95%	23	\$94,344
SouthEast	1017	45,589	6/23/2008	4	4,225	4 No	Biking	Mobile	22%	28,169	Ship	Stand Alone	27	Operations	No	80%	22	\$106,893
SouthEast	1018	69,461	3/19/2008	14	9,016	5 No	Biking	Mobile	39%	9,619	Ship	Stand Alone	29	Operations	No	78%	11	\$228,105
SouthEast	1019	18,647	11/16/2012	6	4,343	5 No	Football	Mobile	37%	36,756	Ship	Mall	20	Operations	No	89%	32	\$109,878
SouthEast	1020	58,424	5/1/2011	5	4,987	3 Yes	Biking	Mobile	52%	17,701	Ship	Mall	25	Operations	Yes	82%	17	\$126,171
SouthEast	1021	10,108	10/4/2012	6	8,855	6 No	Football	PC	40%	36,597	Ship	Mall	24	Operations	Yes	85%	15	\$224,032
SouthEast	1022	46,229	2/17/2014	9	2,260	4 No	Football	PC	35%	48,790	Ship	Mall	40	Operations	Yes	79%	9	\$57,178
SouthEast	1023	20,786	9/27/2014	8	5,074	1 No	Soccer	PC	10%	41,381	No Ship	Mall	22	Operations	Yes	88%	22	\$128,372
SouthEast	1024	12,079	2/8/2012	14	5,314	4 No	Soccer	PC	26%	1	No Ship	Mall	25	Marketing	No	70%	15	\$134,444
SouthEast	1025	61,598	7/4/2012	14	5,309	5 No	Soccer	PC	64%	49,976	No Ship	Mall	26	Marketing	No	65%	16	\$134,318
SouthEast	1026	34,970	1/1/2009	16	3,114	5 No	Soccer	PC	31%	31,897	No Ship	Stand Alone	24	Operations	No	79%	26	\$78,784
SouthEast	1027	29,468	1/22/2011	4	10,315	4 No	Soccer	Mobile	47%	24,234	No Ship	Stand Alone	55	Operations	No	100%	18	\$260,970
SouthEast	1028	42,745	6/21/2009	2	7,196	2 No	Soccer	Mobile	39%	44,774	No Ship	Stand Alone	27	Operations	Yes	99%	6	\$182,059
SouthEast	1029	29,776	4/30/2011	18	4,210	4 No	Soccer	Mobile	17%	41,382	No Ship	Stand Alone	28	Operations	Yes	83%	10	\$106,513
SouthEast	1030	68,304	6/16/2009	12	10,009	2 No	Soccer	Mobile	63%	17,904	No Ship	Stand Alone	19	Operations	Yes	87%	33	\$253,228
SouthEast	1031	72,902	1/18/2011	4	7,053	4 No	Soccer	Mobile	63%	34,176	No Ship	Mall	22	Marketing	Yes	94%	23	\$178,441
SouthEast	1032	70,954	11/29/2008	9	5,914	3 No	Soccer	Mobile	45%	40,330	No Ship	Mall	24	Customer Service	Yes	98%	28	\$149,624
SouthEast	1033	52,637	10/5/2014	10	1,328	1 Yes	Soccer	Mobile	9%	16,420	No Ship	Mall	19	Customer Service	Yes	84%	33	\$33,598
SouthEast	1034	43,199	4/27/2011	4	9,006	2 Yes	Soccer	Mobile	15%	15,031	No Ship	Mall	22	Customer Service	Yes	87%	33	\$227,852
SouthEast	1035	48,716	1/28/2013	4	4,131	2 Yes	Tennis	Mobile	56%	14,533	No Ship	Mall	27	Customer Service	Yes	97%	26	\$104,514
SouthEast	1036	63,830	10/24/2010	5	9,466	2 No	Tennis	Mobile	52%	37,353	No Ship	Mall	34	Customer Service	Yes	81%	14	\$239,490
SouthEast	1037	95,657	5/8/2010	18	8,459	3 No	Tennis	PC	7%	43,471	Ship	Stand Alone	23	Customer Service	No	81%	35	\$214,013
SouthEast	1053	5,441	10/22/2014	17	9,057	4 Yes	Running	PC	37%	23,209	No Ship	Mall	22	Marketing	No	80%	35	\$229,142
SouthEast	1054	5,019	6/1/2013	3	3,442	4 No	WorkOut	Mobile	42%	41,537	No Ship	Mall	28	Marketing	No	89%	11	\$87,083
SouthEast	1055	93,730	12/26/2009	12	5,771	4 No	WorkOut	Mobile	27%	49,872	No Ship	Mall	30	Operations	No	88%	26	\$146,006
SouthEast	1056	58,426	2/25/2014	9	2,502	1 Yes	WorkOut	Mobile	22%	29,010	No Ship	Mall	20	Operations	No	93%	5	\$63,301
SouthEast	1057	70,636	4/19/2008	9	6,190	2 No	Football	Mobile	60%	18,160	No Ship	Stand Alone	27	Operations	No	75%	9	\$156,607
SouthEast	1058	4,777	11/23/2014	11	1,677	4 No	Football	PC	1%	46,171	No Ship	Mall	22	Operations	No	90%	5	\$42,428
SouthEast	1059	101,537	11/8/2013	13	10,439	3 No	Football	PC	8%	15,199	No Ship	Mall	25	Operations	No	97%	16	\$264,107
SouthEast	1060	87,932	11/13/2008	2	3,971	3 No	Soccer	PC	22%	35,540	No Ship	Stand Alone	29	Operations	Yes	92%	8	\$100,466
SouthEast	1061	27,742	6/24/2013	4	2,879	3 No	Tennis	PC	53%	29,051	No Ship	Stand Alone	27	Operations	Yes	89%	9	\$72,839
SouthEast	1062	45,133	10/18/2012	14	1,824	2 No	Tennis	PC	46%	9,102	No Ship	Stand Alone	22	Operations	No	95%	7	\$46,147
SouthEast	1063	65,292	12/25/2008	18	8,263	3 No	Tennis	PC	37%	33,097	No Ship	Mall	28	Operations	Yes	78%	23	\$209,054
SouthEast	1064	76,267	6/6/2011	4	9,839	1 No	Tennis	Mobile	42%	27,848	Ship	Mall	18	Marketing	No	82%	18	\$248,927
SouthEast	1065	86,726	12/7/2009	8	7,862	3 No	WorkOut	Mobile	52%	46,206	Ship	Mall	28	Marketing	No	87%	20	\$198,909
SouthEast	1066	27,906	10/7/2014	12	6,741	4 No	WorkOut	Mobile	33%	21,795	No Ship	Mall	26	Operations	Yes	82%	29	\$170,547
SouthEast	1067	58,056	11/6/2014	5	3,038	4 No	WorkOut	Mobile	10%	28,062	No Ship	Stand Alone	21	Operations	Yes	96%	22	\$76,861
SouthEast	1068	25,952	8/17/2014	15	4,756	1 No	WorkOut	Mobile	34%	11,824	No Ship	Stand Alone	29	Operations	Yes	79%	8	\$120,327
SouthEast	1134	51,299	4/20/2008	15	4,894	4 No	Baseball	Mobile	31%	15,336	Ship	Mall	20	Operations	No	96%	18	\$123,818
SouthEast	1135	86,219	11/20/2014	17	4,418	1 Yes	Baseball	Mobile	3%	47,007	Ship	Stand Alone	27	Operations	Yes	109%	26	\$111,775
SouthEast	1136	74,813	12/5/2014	16	2,899	2 No	WorkOut	Mobile	2%	38,176	Ship	Mall	26	Operations	Yes	99%	26	\$73,345
SouthEast	1137	40,621	6/12/2009	14	8,432	5 No	WorkOut	PC	31%	14,956	Ship	Mall	26	Operations	Yes	78%	12	\$213,330
SouthEast	1138	93,210	5/15/2014	7	1,656	3 Yes	WorkOut	PC	18%	38,109	No Ship	Mall	24	Operations	Yes	87%	20	\$41,897
SouthEast	1139	17,818	10/22/2013	9	7,341	4 Yes	WorkOut	PC	51%	12,524	No Ship	Mall	21	Customer Service	Yes	84%	26	\$185,727
SouthEast	1140	42,540	7/27/2014	4	4,584	1 No	Baseball	PC	-20%	32,344	No Ship	Mall	29	Customer Service	Yes	89%	5	\$115,975
SouthEast	1141	9,545	12/13/2011	12	7,915	1 No	Baseball	PC	50%	11,306	No Ship	Mall	24	Customer Service	Yes	78%	20	\$200,250
SouthEast	1142	89,746	4/29/2008	18	3,653	2 No	Baseball	PC	48%	11,948	No Ship	Mall	43	Customer Service	Yes	93%	17	\$92,421
SouthEast	1143	5,368	3/16/2012	16	7,574	3 No	Baseball	Mobile	6%	38,899	No Ship	Mall	61	Customer Service	Yes	97%	10	\$191,622
SouthEast	1144	76,103	5/11/2012	5	10,215	2 No	Running	Mobile	11%	24,675	No Ship	Mall	22	Customer Service	Yes	99%	30	\$258,440
SouthEast	1145	20,417	9/13/2011	4	2,242	5 Yes	Baseball	Mobile	28%	48,649	Ship	Mall	26	Customer Service	Yes	92%	12	\$56,723
SouthEast	1146	53,031	3/30/2014	9	6,081	5 Yes	Running	Mobile	27%	25,226	Ship	Mall	24	Customer Service	Yes	87%	27	\$153,849
SouthEast	1147	83,985	6/17/2009	17	3,243	4 Yes	Running	Mobile	15%	30,893	Ship	Stand Alone	26	Customer Service	Yes	93%	29	\$42,800,900
SouthEast	1148	55,203	1/21/2014	2	4,017	7 No	Baseball	Mobile	30%	28,936	Ship	Mall	29	Customer Service	Yes	95%	32	\$101,630
SouthEast	1149	3,861	11/1/2010	18	3,469	3 No	Baseball	Mobile	3%	18,294	Ship	Mall	25	Customer Service	Yes	90%	10	\$87,766
SouthEast	1150	21,239	4/16/2009	8	3,961	5 No	Baseball	Mobile	14%	12,917	No Ship	Mall	30	Customer Service	Yes	83%	33	\$100,213
SouthEast	1151	56,580	7/21/2014	11	9,598	5 No	Running	Mobile	18%	34,152	Ship	Mall	26	Customer Service	Yes	77%	7	\$242,829
SouthEast	1152	90,226	3/6/2014	6	2,034	1 N Yes	Running	Mobile	22%	47,556	Ship	Mall	38	Operations	Yes	95%	28	\$51,460
SouthEast	1153	39,070	9/2/2014	14	3,343	1 Yes	Running	Mobile	44%	21,252	Ship	Mall	27	Operations	Yes	75%	31	\$84,578
SouthEast	1154	49,999	11/29/2012	14	1,405	3 No	Running	Mobile	65%	41,046	Ship	Mall	29	Operations	Yes	95%	24	\$35,547
SouthEast	1155	81,785	7/4/2009	5	5,003	5 No	Running	Mobile	50%	18,317	Ship	Mall	105	Operations	No	85%	20	\$126,576
SouthEast	1156	60,310	9/21/2012	4	6,816	5 No	Running	Mobile	38%	35,711	Ship	Mall	29	Operations	No	80%	11	\$172,445

Some of the anomalies or noise in the above data are:

- Under the variable "AI System", the acceptable values are "Yes" and "No". But there are some noise data like "j" and "Nyes" which are wrong inputs and will result in improper calculations and results.
- Similarly, in the column "App", the acceptable values lie in the range "0-100%" as it depicts the percentage of customers who use the store app. But there is a value in the column which reads as "-20%", which is mathematically wrong. This would also result in improper results when we apply data mining techniques.

The yellow boxes in the image above represent these anomalies or mistakes in the data.

2. In the second instance, we identify some **logical errors** in the data. These types of errors may not be corrected by the computer but by human intervention, they can be detected.

Some of the logical mistakes are:

- Under the column "Review", the acceptable values as defined in the question lie between 1-5. But there are some values such as "6" or "7", which lie outside the constraints of the definition of the variable.
- Similarly, under the column "Staff Age", any numerical value can be acceptable. But a person of Age "105" is wrong because a person of that age cannot work in a Sporting Goods store, which involves various physical activities which a person of that age would find extremely difficult to perform.

The gold boxes in the image below represents the logical errors.

Region	Store ID	Population	Store Opening	Sales	Purchase	Coupon	Avg	Sales	% Sales								
				Staff/1000 Sq ft	Monthly Traffic	Review	AI System	App	Facebook	Ship to Store	Store Location	Staff Age	BackGround	Staff College	Parking Places	Total Sales	
SouthEast	1013	62,684	7/13/2012	20	8,202	4 Yes	WorkOut	PC	1%	44,506	Ship	Mall	22 Marketing	Yes	100%	24	\$207,511
SouthEast	1014	96,631	7/10/2010	14	5,628	1 j	WorkOut	PC	50%	47,927	Ship	Stand Alone	27 Marketing	No	105%	14	\$142,388
SouthEast	1015	62,484	1/26/2009	18	3,100	3 No	Biking	PC	18%	47,248	Ship	Mall	28 Marketing	Yes	84%	12	\$78,430
SouthEast	1016	40,108	12/27/2012	14	3,729	1 Yes	Biking	PC	16%	12	Ship	Stand Alone	25 Operations	Yes	95%	23	\$94,344
SouthEast	1017	45,589	6/23/2008	4	4,225	4 No	Biking	Mobile	22%	28,169	Ship	Stand Alone	27 Operations	No	80%	22	\$106,893
SouthEast	1018	69,461	3/19/2008	14	9,016	5 No	Biking	Mobile	39%	9,619	Ship	Stand Alone	29 Operations	No	78%	11	\$228,105
SouthEast	1019	18,647	11/16/2012	6	4,343	5 No	Football	Mobile	37%	36,756	Ship	Mall	20 Operations	No	89%	32	\$109,878
SouthEast	1020	58,424	5/1/2011	5	4,987	3 Yes	Biking	Mobile	52%	17,701	Ship	Mall	25 Operations	Yes	82%	17	\$126,171
SouthEast	1021	10,108	10/4/2012	6	8,855	6 No	Football	PC	40%	36,597	Ship	Mall	24 Operations	Yes	85%	15	\$224,032
SouthEast	1022	46,229	2/17/2014	9	2,260	4 No	Football	PC	35%	48,790	Ship	Mall	40 Operations	Yes	79%	9	\$57,178
SouthEast	1023	20,786	9/27/2014	8	5,074	1 No	Soccer	PC	10%	41,381	No Ship	Mall	22 Operations	Yes	88%	22	\$128,372
SouthEast	1024	12,079	2/8/2012	14	5,314	4 No	Soccer	PC	26%	1.x	No Ship	Mall	25 Marketing	No	70%	15	\$134,444
SouthEast	1025	61,598	7/4/2012	14	5,309	5 No	Soccer	PC	64%	49,976	No Ship	Mall	26 Marketing	No	65%	16	\$134,318
SouthEast	1026	34,970	1/12/2009	16	3,114	5 No	Soccer	PC	31%	31,897	No Ship	Stand Alone	24 Operations	No	79%	26	\$78,784
SouthEast	1027	29,468	1/22/2011	4	10,315	4 No	Soccer	Mobile	47%	24,234	No Ship	Stand Alone	55 Operations	No	100%	18	\$260,970
SouthEast	1028	42,745	6/21/2009	2	7,196	2 No	Soccer	Mobile	39%	44,774	No Ship	Stand Alone	27 Operations	Yes	99%	6	\$182,059
SouthEast	1029	29,776	4/30/2011	18	4,210	4 No	Soccer	Mobile	17%	41,382	No Ship	Stand Alone	28 Operations	Yes	83%	10	\$106,513
SouthEast	1030	68,304	6/16/2009	12	10,009	2 No	Soccer	Mobile	63%	17,904	No Ship	Stand Alone	19 Operations	Yes	87%	33	\$253,228
SouthEast	1031	72,902	1/18/2011	4	7,053	4 No	Soccer	Mobile	63%	34,176	No Ship	Mall	22 Marketing	Yes	94%	23	\$178,441
SouthEast	1032	70,954	11/29/2008	9	5,914	3 No	Soccer	Mobile	45%	40,330	No Ship	Mall	24 Customer Sen	Yes	98%	28	\$149,624
SouthEast	1033	52,637	10/5/2014	10	1,328	1 Yes	Soccer	Mobile	9%	16,420	No Ship	Mall	19 Customer Sen	Yes	84%	33	\$33,598
SouthEast	1034	43,199	4/27/2011	4	9,006	2 Yes	Soccer	Mobile	15%	15,031	No Ship	Mall	22 Customer Sen	Yes	87%	33	\$227,852
SouthEast	1035	48,716	1/28/2013	4	4,131	2 Yes	Tennis	Mobile	56%	14,533	No Ship	Mall	27 Customer Sen	Yes	97%	26	\$104,514
SouthEast	1036	63,830	10/24/2010	5	9,466	2 No	Tennis	Mobile	52%	37,353	No Ship	Mall	34 Customer Sen	Yes	81%	14	\$239,490
SouthEast	1037	95,657	5/8/2010	18	8,459	3 No	Tennis	PC	7%	43,471	Ship	Stand Alone	23 Customer Sen	No	81%	35	\$214,013
SouthEast	1053	5,441	10/22/2014	17	9,057	4 Yes	Running	PC	37%	23,209	No Ship	Mall	22 Marketing	No	80%	35	\$229,142
SouthEast	1054	5,019	6/1/2013	3	3,442	4 No	WorkOut	Mobile	42%	41,537	No Ship	Mall	28 Marketing	No	89%	11	\$87,083
SouthEast	1055	93,730	12/26/2009	12	5,771	4 No	WorkOut	Mobile	27%	49,872	No Ship	Mall	30 Operations	No	88%	26	\$146,006
SouthEast	1056	58,426	2/25/2014	9	2,502	1 Yes	WorkOut	Mobile	22%	29,010	No Ship	Mall	20 Operations	No	93%	5	\$63,301
SouthEast	1057	70,636	4/19/2008	9	6,190	2 No	Football	Mobile	60%	18,160	No Ship	Stand Alone	27 Operations	No	75%	9	\$156,607
SouthEast	1058	4,777	11/23/2014	11	1,677	4 No	Football	PC	1%	46,171	No Ship	Mall	22 Operations	No	90%	5	\$42,428
SouthEast	1059	101,537	11/8/2013	13	10,439	3 No	Football	PC	8%	15,199	No Ship	Mall	25 Operations	No	97%	16	\$264,107
SouthEast	1060	87,932	11/13/2008	2	3,971	3 No	Soccer	PC	22%	35,540	No Ship	Stand Alone	29 Operations	Yes	92%	8	\$100,466
SouthEast	1061	27,742	6/24/2013	4	2,879	3 No	Tennis	PC	53%	29,051	No Ship	Stand Alone	27 Operations	Yes	89%	9	\$72,839
SouthEast	1062	45,133	10/18/2012	14	1,824	2 No	Tennis	PC	46%	9,102	No Ship	Stand Alone	22 Operations	No	95%	7	\$46,147
SouthEast	1063	65,292	12/25/2008	18	8,263	3 j	Tennis	PC	37%	33,097	No Ship	Mall	28 Operations	Yes	78%	23	\$209,054
SouthEast	1064	76,267	6/6/2011	4	9,839	1 No	Tennis	Mobile	42%	27,848	Ship	Mall	18 Marketing	No	82%	18	\$248,927
SouthEast	1065	86,726	12/7/2009	8	7,862	3 No	WorkOut	Mobile	52%	46,206	Ship	Mall	28 Marketing	No	87%	20	\$198,909
SouthEast	1066	27,906	10/7/2014	12	6,741	4 No	WorkOut	Mobile	33%	21,795	No Ship	Mall	26 Operations	Yes	82%	29	\$170,547
SouthEast	1067	58,056	11/6/2014	5	3,038	4 No	WorkOut	Mobile	10%	28,062	No Ship	Stand Alone	21 Operations	Yes	96%	22	\$76,861
SouthEast	1068	25,952	8/17/2014	15	4,756	1 No	WorkOut	Mobile	34%	11,824	No Ship	Stand Alone	29 Operations	Yes	79%	8	\$120,327
SouthEast	1134	51,299	4/20/2008	15	4,894	4 No	Baseball	Mobile	31%	15,336	Ship	Mall	20 Operations	No	96%	18	\$123,818
SouthEast	1135	86,219	11/20/2014	17	4,418	1 Yes	Baseball	Mobile	3%	47,007	Ship	Stand Alone	27 Operations	Yes	109%	26	\$111,775
SouthEast	1136	74,813	12/5/2014	16	2,899	2 No	WorkOut	Mobile	2%	38,176	Ship	Mall	26 Operations	Yes	99%	26	\$73,345
SouthEast	1137	40,621	6/12/2009	14	8,432	5 No	WorkOut	PC	31%	14,956	Ship	Mall	26 Operations	Yes	78%	12	\$213,330
SouthEast	1138	93,210	5/15/2014	7	1,656	3 Yes	WorkOut	PC	18%	38,109	No Ship	Mall	24 Operations	Yes	87%	20	\$41,897
SouthEast	1139	17,818	10/22/2013	9	7,341	4 Yes	WorkOut	PC	51%	12,524	No Ship	Mall	21 Customer Sen	Yes	84%	26	\$185,727
SouthEast	1140	42,540	7/27/2014	4	4,584	1 No	Baseball	PC	-20%	32,344	No Ship	Mall	29 Customer Sen	Yes	89%	5	\$115,975
SouthEast	1141	9,545	12/13/2011	12	7,915	1 No	Baseball	PC	50%	11,306	No Ship	Mall	24 Customer Sen	Yes	78%	20	\$200,250
SouthEast	1142	89,746	4/29/2008	18	3,653	2 No	Baseball	PC	48%	11,948	No Ship	Mall	43 Customer Sen	Yes	93%	17	\$92,421
SouthEast	1143	5,368	3/16/2012	16	7,574	3 No	Baseball	Mobile	6%	38,899	No Ship	Mall	61 Customer Sen	Yes	97%	10	\$191,622
SouthEast	1144	76,103	5/11/2012	5	10,215	2 No	Running	Mobile	11%	24,675	No Ship	Mall	22 Customer Sen	Yes	99%	30	\$258,440
SouthEast	1145	20,417	9/13/2011	4	2,242	5 Yes	Baseball	Mobile	28%	48,649	Ship	Mall	26 Customer Sen	Yes	92%	12	\$56,723
SouthEast	1146	53,031	3/30/2014	9	6,081	5 Yes	Running	Mobile	27%	25,226	Ship	Mall	24 Customer Sen	Yes	87%	27	\$153,849
SouthEast	1147	83,985	6/17/2009	17	3,243	4 Yes	Running	Mobile	15%	30,893	Ship	Stand Alone	26 Customer Sen	Yes	93%	29	\$42,800,900
SouthEast	1148	55,203	1/21/2014	2	4,017	7 No	Baseball	Mobile	30%	28,936	Ship	Mall	29 Customer Sen	Yes	95%	32	\$101,630
SouthEast	1149	3,861	11/11/2010	18	3,469	3 No	Baseball	Mobile	3%	18,294	Ship	Mall	25 Customer Sen	Yes	90%	10	\$87,766
SouthEast	1150	21,239	4/16/2009	8	3,961	5 No	Baseball	Mobile	14%	12,917	Ship	Mall	30 Customer Sen	Yes	83%	33	\$100,213
SouthEast	1151	56,580	7/21/2014	11	9,598	5 No	Running	Mobile	18%	34,152	Ship	Mall	26 Customer Sen	Yes	77%	7	\$242,829
SouthEast	1152	90,226	3/6/2014	6	2,034	1 Nyes	Running	Mobile	22%	47,556	Ship	Mall	38 Operations	Yes	95%	28	\$51,460
SouthEast	1153	39,070	9/2/2014	14	3,343	1 Yes	Running	Mobile	44%	21,252	Ship	Mall	27 Operations	Yes	75%	31	\$84,578
SouthEast	1154	49,999	11/29/2012	14	1,405	3 No	Running	Mobile	65%	41,046	Ship	Mall	29 Operations	Yes	95%	24	\$35,547
SouthEast	1155	81,785	7/4/2009	5	5,003	5 No	Running	Mobile	50%	18,317	Ship	Mall	105 Operations	No	85%	20	\$126,576
SouthEast	1156	60,310	9/21/2012	4	6,816	5 No	Running	Mobile	38%	35,711	Ship	Mall	29 Operations	No	80%	11	\$172,445

3. In this step, we extract **weekly** data for some of the variables.

Region	Store ID	Population	Store Opening	11/28/2011	No of weeks store open	Sales Staff/1000 Sq ft	Monthly Traffic	Weekly Traffic	Review	AI System	Product Purchase d	Coupon Received	App	Avg Monthly Facebook	Avg Weekly Facebook	Ship to Store	Store Location	Staff Age	Sales BackGround		Loyalty Card	% Sales Staff College	Parking Places	Total Sales
																			Gender	Age Group				
SouthEast	1013	62,684	7/13/2012	3425	489	20	8,202	2,051	4 Yes	WorkOut	PC	1%	44,506	11,127	Ship	Mall	22	Marketing	Yes	100%	24	\$207,511		
SouthEast	1014	96,631	7/10/2010	4159	594	14	5,628	1,407	1 No	WorkOut	PC	50%	47,927	11,982	Ship	Stand Alone	27	Marketing	No	105%	12	\$142,388		
SouthEast	1015	62,464	1/26/2009	4689	670	18	3,100	775	3 No	Biking	PC	18%	47,248	11,812	Ship	Mall	28	Marketing	Yes	84%	12	\$78,430		
SouthEast	1016	40,108	12/27/2012	3258	465	14	3,729	932	1 Yes	Biking	PC	16%	47,12	#VALUE!	Ship	Stand Alone	25	Operations	Yes	95%	23	\$94,344		
SouthEast	1017	45,586	6/23/2008	4906	701	4	4,225	1,056	4 No	Biking	Mobile	22%	28,169	7,042	Ship	Stand Alone	27	Operations	No	80%	22	\$106,893		
SouthEast	1018	69,461	3/19/2008	5002	715	14	9,016	2,254	5 No	Biking	Mobile	39%	9,619	2,405	Ship	Stand Alone	29	Operations	No	78%	11	\$228,105		
SouthEast	1019	18,647	11/16/2012	3299	471	6	4,343	1,086	5 No	Football	Mobile	37%	36,756	9,189	Ship	Mall	20	Operations	No	89%	32	\$109,878		
SouthEast	1020	58,424	5/1/2011	3864	552	5	4,987	1,247	3 Yes	Biking	Mobile	52%	17,701	4,425	Ship	Mall	25	Operations	Yes	82%	17	\$126,171		
SouthEast	1021	10,108	10/4/2012	3342	477	6	8,855	2,214	6 No	Football	PC	40%	36,597	9,149	Ship	Mall	24	Operations	Yes	85%	15	\$224,032		
SouthEast	1022	46,229	2/17/2014	2841	408	9	2,280	565	4 No	Football	PC	35%	48,790	12,198	Ship	Mall	40	Operations	Yes	79%	9	\$57,178		
SouthEast	1023	20,786	9/27/2014	2619	374	8	5,074	1,269	1 No	Soccer	PC	10%	41,381	10,345	Ship	Mall	22	Operations	Yes	88%	22	\$128,372		
SouthEast	1024	12,079	2/8/2012	3581	512	14	5,314	1,329	4 No	Soccer	PC	26%	1,xi	No Ship	Mall	25	Marketing	No	70%	15	\$134,444			
SouthEast	1025	61,598	7/4/2012	3434	491	14	5,309	1,327	5 No	Soccer	PC	64%	49,976	12,494	No Ship	Mall	26	Marketing	No	65%	16	\$134,318		
SouthEast	1026	34,970	1/12/2009	4703	672	16	3,114	779	5 No	Soccer	PC	31%	31,897	7,974	No Ship	Stand Alone	24	Operations	No	79%	26	\$78,784		
SouthEast	1027	29,468	1/22/2011	3963	566	4	10,315	2,579	4 No	Soccer	Mobile	47%	24,234	6,059	No Ship	Stand Alone	55	Operations	No	100%	18	\$260,970		
SouthEast	1028	42,745	6/21/2009	4543	649	2	7,196	1,799	2 No	Soccer	Mobile	39%	44,774	11,194	No Ship	Stand Alone	27	Operations	Yes	99%	6	\$182,059		
SouthEast	1029	29,776	4/30/2011	3865	552	18	4,210	1,053	4 No	Soccer	Mobile	17%	41,382	10,346	No Ship	Stand Alone	28	Operations	Yes	83%	10	\$106,513		
SouthEast	1030	68,304	6/16/2009	4548	650	12	10,009	2,902	2 No	Soccer	Mobile	63%	17,904	4,476	No Ship	Stand Alone	19	Operations	Yes	87%	33	\$255,228		
SouthEast	1031	72,902	1/18/2011	3967	567	4	7,053	1,763	4 No	Soccer	Mobile	63%	34,176	8,544	No Ship	Mall	22	Marketing	Yes	94%	23	\$178,441		
SouthEast	1032	70,954	11/29/2008	4747	678	9	5,914	1,479	3 No	Soccer	Mobile	45%	40,330	10,083	No Ship	Mall	24	Customer S Yes	98%	28	\$149,624			
SouthEast	1033	52,637	10/5/2014	2611	373	10	1,328	332	1 Yes	Soccer	Mobile	9%	16,420	4,105	No Ship	Mall	19	Customer S Yes	84%	33	\$33,598			
SouthEast	1034	43,199	4/27/2011	3968	553	4	9,006	2,252	2 Yes	Soccer	Mobile	15%	15,031	3,758	No Ship	Mall	22	Customer S Yes	87%	33	\$227,852			
SouthEast	1035	48,716	1/28/2013	3226	461	4	4,131	1,033	2 Yes	Tennis	Mobile	56%	14,533	3,633	No Ship	Mall	27	Customer S Yes	97%	26	\$104,514			
SouthEast	1036	63,830	10/24/2010	4053	579	5	9,466	2,367	2 No	Tennis	Mobile	52%	37,353	9,338	No Ship	Mall	34	Customer S Yes	81%	14	\$239,490			
SouthEast	1037	95,657	5/8/2010	4222	603	18	8,459	2,115	3 No	Tennis	PC	7%	43,471	10,868	Ship	Stand Alone	23	Customer S No	81%	35	\$214,013			
SouthEast	1038	5,441	10/22/2014	2594	371	17	9,057	2,264	4 Yes	Running	PC	37%	23,209	5,602	No Ship	Mall	22	Marketing	No	80%	35	\$229,142		
SouthEast	1039	5,019	6/1/2013	3102	443	3	3,442	861	4 No	WorkOut	Mobile	42%	41,537	10,384	No Ship	Mall	28	Marketing	No	89%	11	\$87,083		
SouthEast	1040	93,730	12/26/2009	4355	622	12	5,771	1,443	4 No	WorkOut	Mobile	27%	49,672	12,468	No Ship	Mall	30	Operations	No	88%	26	\$146,006		
SouthEast	1041	58,426	2/25/2014	2833	405	9	2,502	626	1 Yes	WorkOut	Mobile	22%	29,010	7,253	No Ship	Mall	20	Operations	No	93%	5	\$63,301		
SouthEast	1042	4,777	11/23/2014	2562	366	11	1,677	419	4 No	Football	PC	1%	46,171	11,543	No Ship	Mall	22	Operations	No	90%	5	\$42,428		
SouthEast	1043	101,537	11/8/2013	2942	420	13	10,439	2,610	3 No	Football	PC	8%	15,199	3,800	No Ship	Mall	25	Operations	No	97%	16	\$264,107		
SouthEast	1044	87,932	11/13/2008	4763	680	2	3,971	993	3 No	Soccer	PC	22%	35,540	8,685	No Ship	Stand Alone	29	Operations	Yes	92%	8	\$100,466		
SouthEast	1045	27,742	6/24/2013	3079	440	4	2,879	720	3 No	Tennis	PC	53%	29,051	7,263	No Ship	Stand Alone	27	Operations	Yes	89%	9	\$72,839		
SouthEast	1046	45,133	10/18/2012	3328	475	14	1,824	456	2 No	Tennis	PC	46%	9,102	2,276	No Ship	Stand Alone	22	Operations	No	95%	7	\$46,147		
SouthEast	1047	65,292	12/25/2008	4721	674	18	8,263	2,066	3 No	Tennis	PC	37%	33,097	8,274	No Ship	Mall	28	Operations	Yes	78%	23	\$209,054		
SouthEast	1048	76,267	6/6/2011	3828	547	4	9,839	2,460	1 No	Tennis	Mobile	42%	27,848	6,962	Ship	Mall	18	Marketing	No	82%	18	\$248,927		
SouthEast	1049	86,726	12/7/2009	4374	625	8	7,862	1,966	3 No	WorkOut	Mobile	52%	46,206	11,552	Ship	Mall	28	Marketing	No	87%	20	\$198,909		
SouthEast	1050	27,906	10/7/2014	2605	373	12	6,741	1,685	4 No	WorkOut	Mobile	33%	21,795	5,449	Ship	Mall	26	Operations	Yes	82%	29	\$170,547		
SouthEast	1051	58,056	11/6/2014	2579	368	5	3,038	760	4 No	WorkOut	Mobile	10%	28,062	7,016	No Ship	Stand Alone	21	Operations	Yes	96%	22	\$76,861		
SouthEast	1052	25,952	8/17/2014	2660	380	15	4,756	1,169	1 No	WorkOut	Mobile	34%	11,824	2,950	No Ship	Stand Alone	29	Operations	Yes	79%	8	\$120,327		
SouthEast	1053	51,299	4/20/2008	4970	710	15	4,894	1,224	4 No	Baseball	Mobile	31%	15,336	3,834	Ship	Mall	20	Operations	No	96%	18	\$123,818		
SouthEast	1054	86,219	11/20/2014	2565	366	17	4,418	1,105	1 Yes	Baseball	Mobile	3%	47,007	11,752	Ship	Stand Alone	27	Operations	Yes	109%	26	\$111,775		
SouthEast	1055	9,145	12/5/2014	2550	364	16	2,899	725	2 No	WorkOut	Mobile	2%	38,176	9,544	Ship	Mall	26	Operations	Yes	99%	26	\$73,345		
SouthEast	1056	137	6/1/2009	4552	650	14	8,432	2,108	5 No	WorkOut	PC	31%	14,956	3,739	Ship	Mall	26	Operations	Yes	78%	12	\$213,330		
SouthEast	1057	93,210	5/15/2014	2754	393	7	1,656	414	3 Yes	WorkOut	PC	18%	38,109	9,527	No Ship	Mall	24	Operations	Yes	87%	20	\$41,897		
SouthEast	1058	17,818	10/22/2013	2955	423	9	7,341	1,835	4 Yes	WorkOut	PC	51%	12,524	3,131	No Ship	Mall	21	Customer S Yes	84%	26	\$185,727			
SouthEast	1059	42,540	7/27/2014	2681	383	4	4,584	1,146	1 No	Baseball	PC	-20%	32,344	8,080	No Ship	Mall	29	Customer S Yes	89%	5	\$115,975			
SouthEast	1060	9,545	12/13/2011	3638	520	12	7,915	1,979	1 No	Baseball	PC	50%	11,306	2,827	No Ship	Mall	24	Customer S Yes	78%	20	\$200,250			
SouthEast	1061	89,746	4/29/2008	4961	709	18	3,653	913	2 No	Baseball	PC	48%	11,948	2,987	No Ship	Mall	43	Customer S Yes	93%	17	\$92,421			
SouthEast	1062	5,368	3/16/2012	3544	506	16	7,574	1,894	3 No	Baseball	Mobile	6%	38,899	9,725	No Ship	Mall	61	Customer S Yes	97%	10	\$191,622			
SouthEast	1063	76,103	5/11/2012	3488	498	5	10,215	2,554	2 No	Running	Mobile	11%	24,675	6,169	No Ship	Mall	22	Customer S Yes	99%	30	\$258,440			
SouthEast	1064	20,417	9/13/2011	3729	533	4	2,242	561	5 Yes	Baseball	Mobile	20%	48,649	12,162	Ship	Mall	26	Customer S Yes	92%	12	\$56,723			
SouthEast	1065	53,031	3/30/2014	2800	400	9	6,081	1,520	5 Yes	Running	Mobile	27%	25,226	6,307	Ship	Mall	24	Customer S Yes	87%	27	\$153,849			
SouthEast	1066	83,985	6/17/2009	4547	650	17	3,243																	

4. In this step, we identify all the redundant variables and remove all the errors.

Region	Store ID	Population n	Store Opening	11/28/2011	No of weeks Store open	Sales Staff/1000 Sq Ft	Monthly Traffic	Weekly Traffic	Review 4 Yes	AI System	Product Purchase d	Coupon Receive	App	Avg Monthly Facebook	Avg Weekly Facebook	Ship to Store	Store Location	Staff Age	Sales BackGround	Loyalty Card	% Sales Staff College	Parking Places	Total Sales
SouthEast	1013	62,684	7/13/2012	3425	489	20	8,202	2,051						44,506	11,127	Ship Mall	22 Marketing	Yes	100%	24	\$207,511		
SouthEast	1014	96,531	7/10/2010	4159	594	14	5,628	1,407	1 x	WorkOut	PC	50%	47,927	11,992	Ship Stand Alone	27 Marketing	No	105%	16	\$142,398			
SouthEast	1015	62,484	1/26/2009	4699	670	18	3,100	775			3 No	Biking	PC	18%	47,248	7,042	Ship Mall	28 Marketing	Yes	94%	12	\$78,430	
SouthEast	1016	83,392	12/27/2012	3258	485	14	3,729	932	1 x	Biking	PC	10%	47,927	12,412	Ship Stand Alone	29 Operations	Yes	95%	23	\$60,334			
SouthEast	1017	45,589	6/23/2008	4906	701	4	4,225	1,056			4 No	Biking	Mobile	22%	28,169	7,042	Stand Alone	27 Operations	No	89%	22	\$105,693	
SouthEast	1018	69,461	3/19/2008	5002	715	14	9,016	2,254			5 No	Biking	Mobile	39%	9,619	2,405	Ship Stand Alone	29 Operations	No	78%	11	\$228,105	
SouthEast	1019	18,647	11/16/2012	3269	471	8	4,343	1,086			5 No	Football	Mobile	37%	36,756	9,169	Ship Mall	20 Operations	No	89%	32	\$109,878	
SouthEast	1020	58,424	5/1/2011	3964	552	5	4,987	1,247			3 Yes	Biking	Mobile	52%	17,701	4,425	Ship Mall	25 Operations	Yes	82%	17	\$126,171	
SouthEast	1021	53,168	10/4/2012	3342	477	6	8,655	2,214			3 No	Biking	PC	40%	36,597	9,149	Ship Mall	29 Operations	Yes	85%	10	\$229,078	
SouthEast	1022	46,229	2/17/2014	2841	406	9	2,280	565			4 No	Football	PC	35%	46,790	12,198	Ship Mall	40 Operations	Yes	79%	9	\$57,178	
SouthEast	1023	20,786	9/27/2014	2619	374	8	5,074	1,269			1 No	Soccer	PC	10%	41,381	10,345	No Ship Mall	22 Operations	Yes	88%	22	\$129,372	
SouthEast	1024	12,079	2/8/2012	3581	512	14	5,314	1,329			4 No	Soccer	PC	29%	1 x	No Ship	Mall	29 Marketing	No	70%	15	\$134,444	
SouthEast	1025	61,598	7/4/2012	3424	491	14	5,309	1,327			5 No	Soccer	PC	64%	49,976	12,494	No Ship Mall	26 Marketing	No	65%	16	\$134,318	
SouthEast	1026	34,970	1/12/2009	4703	672	16	3,114	779			5 No	Soccer	PC	31%	31,897	7,974	No Ship Stand Alone	24 Operations	No	79%	26	\$78,784	
SouthEast	1027	29,468	1/22/2011	3963	566	4	10,315	2,579			4 No	Soccer	Mobile	47%	24,234	6,059	No Ship Stand Alone	55 Operations	No	100%	18	\$260,970	
SouthEast	1028	42,745	6/21/2009	4543	649	2	7,195	1,799			2 No	Soccer	Mobile	39%	44,774	11,194	No Ship Stand Alone	27 Operations	Yes	99%	6	\$182,059	
SouthEast	1029	29,776	4/30/2010	3965	552	18	4,210	1,053			4 No	Soccer	Mobile	17%	41,382	10,346	No Ship Stand Alone	28 Operations	Yes	83%	10	\$105,513	
SouthEast	1030	68,304	6/16/2009	4548	650	12	10,000	2,502			2 No	Soccer	Mobile	63%	17,804	4,476	No Ship Stand Alone	19 Operations	Yes	87%	33	\$253,228	
SouthEast	1031	72,902	1/18/2011	3967	567	4	7,053	1,763			4 No	Soccer	Mobile	63%	34,176	8,544	No Ship Mall	22 Marketing	Yes	94%	23	\$178,441	
SouthEast	1032	70,954	11/29/2008	4747	678	9	5,914	1,479			3 No	Soccer	Mobile	45%	40,330	10,083	No Ship Mall	24 Customer Servi Yes	98%	28	\$149,624		
SouthEast	1033	52,637	10/5/2014	2611	373	10	1,320	332			1 Yes	Soccer	Mobile	9%	16,420	4,105	No Ship Mall	19 Customer Servi Yes	84%	33	\$33,598		
SouthEast	1034	43,199	4/27/2011	3968	553	4	9,906	2,252			2 Yes	Soccer	Mobile	15%	15,031	3,758	No Ship Mall	22 Customer Servi Yes	87%	33	\$227,852		
SouthEast	1035	48,716	1/28/2009	3230	461	4	4,131	1,033			2 Yes	Tennis	Mobile	56%	14,533	3,633	No Ship Mall	27 Customer Servi Yes	97%	26	\$104,514		
SouthEast	1036	63,830	10/24/2010	4053	579	5	9,966	2,367			2 No	Tennis	Mobile	52%	37,353	9,338	No Ship Mall	34 Customer Servi Yes	81%	14	\$239,490		
SouthEast	1037	95,657	5/9/2010	4222	603	18	8,459	2,115			3 No	Tennis	PC	7%	43,471	10,886	No Ship Stand Alone	23 Customer Servi No	81%	35	\$214,013		
SouthEast	1038	5,441	12/22/2014	2594	371	17	9,057	2,264			4 Yes	Running	PC	37%	23,209	5,802	No Ship Mall	22 Marketing	No	89%	35	\$229,142	
SouthEast	1039	5,019	6/1/2013	3102	443	3	3,442	861			4 No	WorkOut	Mobile	42%	41,537	10,384	No Ship Mall	28 Marketing	No	69%	11	\$87,083	
SouthEast	1040	93,730	12/26/2009	4355	622	12	5,771	1,443			4 No	WorkOut	Mobile	27%	49,872	12,468	No Ship Mall	30 Operations	No	89%	26	\$146,006	
SouthEast	1041	58,426	4/25/2010	2833	405	9	2,542	626			1 Yes	WorkOut	Mobile	22%	29,010	7,253	No Ship Mall	20 Operations	No	93%	5	\$63,301	
SouthEast	1042	70,636	4/19/2009	4971	710	9	6,190	1,548			2 No	Football	Mobile	60%	18,160	4,540	No Ship Stand Alone	27 Operations	No	79%	9	\$156,607	
SouthEast	1043	4,777	11/23/2014	2505	366	11	1,077	419			4 No	Football	PC	1%	46,171	11,543	No Ship Mall	22 Operations	No	90%	5	\$42,428	
SouthEast	1044	101,537	11/8/2013	2945	420	13	10,439	2,610			3 No	Football	PC	8%	15,199	3,800	No Ship Mall	29 Operations	No	97%	16	\$264,107	
SouthEast	1045	87,932	11/13/2008	4763	680	2	3,971	993			3 No	Soccer	PC	22%	35,540	8,885	No Ship Stand Alone	29 Operations	Yes	92%	8	\$100,466	
SouthEast	1046	27,742	6/24/2013	3079	440	4	2,879	720			3 No	Tennis	PC	53%	29,051	7,263	No Ship Stand Alone	27 Operations	Yes	89%	9	\$72,839	
SouthEast	1047	45,133	10/18/2012	3358	475	14	1,324	456			2 No	Tennis	PC	46%	9,102	2,276	No Ship Stand Alone	22 Operations	Yes	95%	7	\$46,147	
SouthEast	1048	65,269	12/25/2008	4721	674	18	8,263	2,066	3 Yes	Tennis	PC	37%	33,097	9,274	No Ship Mall	29 Operations	Yes	70%	23	\$209,064			
SouthEast	1049	76,267	6/4/2011	3829	547	4	9,939	2,460	1 No	Tennis	Mobile	42%	27,848	6,962	Ship Mall	18 Marketing	No	82%	18	\$249,927			
SouthEast	1050	88,726	12/7/2009	4314	625	8	7,082	1,966	3 No	WorkOut	Mobile	52%	46,206	11,552	Ship Mall	28 Marketing	No	67%	20	\$198,909			
SouthEast	1051	27,906	10/7/2014	2609	373	12	6,741	1,685	4 No	WorkOut	Mobile	33%	21,795	5,449	No Ship Mall	26 Operations	Yes	82%	29	\$170,547			
SouthEast	1052	58,056	11/6/2014	2579	368	5	3,039	780	4 No	WorkOut	Mobile	10%	28,062	7,016	No Ship Stand Alone	21 Operations	Yes	96%	22	\$76,861			
SouthEast	1053	25,952	6/17/2014	2699	380	15	4,756	1,189	1 No	WorkOut	Mobile	34%	11,824	2,956	No Ship Stand Alone	29 Operations	Yes	79%	8	\$120,327			
SouthEast	1054	51,299	4/20/2009	4970	710	15	4,894	1,224	4 No	Baseball	Mobile	31%	15,336	3,834	No Ship Mall	20 Operations	No	98%	18	\$123,818			
SouthEast	1055	25,505	11/20/2014	2505	366	17	4,410	1,509			4 No	Baseball	PC	47,007	14,744	No Ship Stand Alone	29 Operations	Yes	109%	23	\$209,064		
SouthEast	1056	74,813	12/5/2014	2550	364	16	2,699	725	2 No	WorkOut	Mobile	2%	38,176	9,544	Ship Mall	26 Operations	Yes	99%	26	\$73,345			
SouthEast	1057	40,621	6/12/2009	4555	650	14	8,430	2,106	5 No	WorkOut	PC	31%	14,956	3,739	Ship Mall	26 Operations	Yes	79%	12	\$213,330			
SouthEast	1058	93,210	5/15/2014	2754	393	7	1,606	414	3 Yes	WorkOut	PC	18%	38,109	9,527	No Ship Mall	24 Operations	Yes	87%	20	\$41,897			
SouthEast	1059	17,818	10/22/2013	2550	423	9	7,341	1,835	4 Yes	WorkOut	PC	51%	12,524	3,131	No Ship Mall	21 Customer Servi Yes	84%	26	\$185,727				
SouthEast	1060	42,540	7/27/2014	2681	383	4	4,584	1,146	1 No	Baseball	PC	20%	32,344	8,088	No Ship Mall	29 Customer Servi Yes	89%	5	\$115,975				
SouthEast	1061	11,441	9,545	12/12/2011	3630	520	12	7,915	1,979	1 No	Baseball	PC	50%	11,308	2,827	No Ship Mall	24 Customer Servi Yes	78%	20	\$200,250			
SouthEast	1062	11,442	89,746	4/29/2008	4961	709	18	3,053	913	2 No	Baseball	PC	48%	11,948	2,987	No Ship Mall	43 Customer Servi Yes	93%	17	\$92,421			
SouthEast	1063	11,443	5,368	3/16/2012	3544	506	16	7,574	1,894	3 No	Baseball	Mobile	6%	38,899	9,725	No Ship Mall	61 Customer Servi Yes	97%	10	\$191,622			
SouthEast	1064	11,444	76,103	5/11/2012	3489	498	5	10,215	2,554	2 No	Running	Mobile	11%	24,675	6,169	No Ship Mall	22 Customer Servi Yes	99%	30	\$258,440			
SouthEast	1065	10,445	21,239	4/16/2009	2611	533	4	2,242	561	5 Yes	Baseball	Mobile	28%	48,649	12,182	Ship Mall	26 Customer Servi Yes	92%	12	\$56,723			
SouthEast	1066	53,031	3/30/2014	2890	400	9	6,081	1,520	5 Yes	Running	Mobile	27%	25,226	6,307	Ship Mall	24 Customer Servi Yes	87%	27	\$153,849				
SouthEast	1067	83,985	6/17/2009	4547	650	17	3,243	811	4 Yes	Running	Mobile	15%	30,893	7,723	Ship Stand Alone	26 Customer Servi Yes	93%	29	\$42,800,900				
SouthEast	1068	11,449	3,861	11/11/2010	4032	576	18	3,469	867	3 No	Baseball	Mobile	3%	18,394	4,574	Ship Mall	25 Customer Servi Yes	90%	10	\$87,766			
SouthEast	1069	11,500	21,239	4/16/2009	4605	658	8	3,961	990	5 No	Baseball	Mobile	14%	12,917	3,229	Ship Mall	30 Customer Servi Yes	83%	33	\$100,213</			

The final table after removing the redundant variables and the anomalies and errors is shown below

Region	Store ID	Population n	No of weeks Store open	Weekly Traffic	Review	AI System	Product Purchase d	Coupon Receive	App	Avg Weekly Facebook	Ship to Store	Store Location	Staff Age	Sales BackGround	Loyalty Card	Parking Places	Total Sales
SouthEast	1013	62,684	489	2,051	4 Yes	WorkOut	PC	1%	11,127	Ship	Mall	22	Marketing	Yes	24	\$207,511	
SouthEast	1015	62,484	670	775	3 No	Biking	PC	18%	11,812	Ship	Mall	28	Marketing	Yes	12	\$78,430	
SouthEast	1017	45,589	701	1,056	4 No	Biking	Mobile	22%	7,042	Ship	Stand Alone	27	Operations	No	22	\$106,893	
SouthEast	1018	69,461	715	2,254	5 No	Biking	Mobile	39%	2,405	Ship	Stand Alone	29	Operations	No	11	\$228,105	
SouthEast	1019	18,647	471	1,086	5 No	Football	Mobile	37%	9,189	Ship	Mall	20	Operations	No	32	\$109,878	
SouthEast	1020	58,424	552	1,247	3 Yes	Biking	Mobile	52%	4,425	Ship	Mall	25	Operations	Yes	17	\$126,171	
SouthEast	1022	46,229	406	565	4 No	Football	PC	35%	12,198	Ship	Mall	40	Operations	Yes	9	\$57,178	
SouthEast	1023	20,786	374	1,269	1 No	Soccer	PC	10%	10,345	No Ship	Mall	22	Operations	Yes	22	\$128,372	
SouthEast	1025	61,598	491	1,327	5 No	Soccer	PC	64%	12,494	No Ship	Mall	26	Marketing	No	16	\$134,318	
SouthEast	1026	34,970	672	779	5 No	Soccer	PC	31%	7,974	No Ship	Stand Alone	24	Operations	No	26	\$78,784	
SouthEast	1027	29,468	566	2,579	4 No	Soccer	Mobile	47%	6,059	No Ship	Stand Alone	55	Operations	No	18	\$260,970	
SouthEast	1028	42,745	649	1,799	2 No	Soccer	Mobile	39%	11,194	No Ship	Stand Alone	27	Operations	Yes	6	\$182,059	
SouthEast	1029	29,776	552	1,053	4 No	Soccer	Mobile	17%	10,346	No Ship	Stand Alone	28	Operations	Yes	10	\$106,513	
SouthEast	1030	68,304	650	2,502	2 No	Soccer	Mobile	63%	4,476	No Ship	Stand Alone	19	Operations	Yes	33	\$253,228	
SouthEast	1031	72,902	567	1,763	4 No	Soccer	Mobile	63%	8,544	No Ship	Mall	22	Marketing	Yes	23	\$178,441	
SouthEast	1032	70,954	678	1,479	3 No	Soccer	Mobile	45%	10,083	No Ship	Mall	24	Customer S	Yes	28	\$149,624	
SouthEast	1033	52,637	373	332	1 Yes	Soccer	Mobile	9%	4,105	No Ship	Mall	19	Customer S	Yes	33	\$33,598	
SouthEast	1034	43,199	553	2,252	2 Yes	Soccer	Mobile	15%	3,758	No Ship	Mall	22	Customer S	Yes	33	\$227,852	
SouthEast	1035	48,716	461	1,033	2 Yes	Tennis	Mobile	56%	3,633	No Ship	Mall	27	Customer S	Yes	26	\$104,514	
SouthEast	1036	63,830	579	2,367	2 No	Tennis	Mobile	52%	9,338	No Ship	Mall	34	Customer S	Yes	14	\$239,490	
SouthEast	1037	95,657	603	2,115	3 No	Tennis	PC	7%	10,868	Ship	Stand Alone	23	Customer S	No	35	\$214,013	
SouthEast	1053	5,441	371	2,264	4 Yes	Running	PC	37%	5,802	No Ship	Mall	22	Marketing	No	35	\$229,142	
SouthEast	1054	5,019	443	861	4 No	WorkOut	Mobile	42%	10,384	No Ship	Mall	28	Marketing	No	11	\$87,083	
SouthEast	1055	93,730	622	1,443	4 No	WorkOut	Mobile	27%	12,468	No Ship	Mall	30	Operations	No	26	\$146,006	
SouthEast	1056	58,426	405	626	1 Yes	WorkOut	Mobile	22%	7,253	No Ship	Mall	20	Operations	No	5	\$63,301	
SouthEast	1057	70,636	710	1,548	2 No	Football	Mobile	60%	4,540	No Ship	Stand Alone	27	Operations	No	9	\$156,607	
SouthEast	1058	4,777	366	419	4 No	Football	PC	1%	11,543	No Ship	Mall	22	Operations	No	5	\$42,428	
SouthEast	1059	101,537	420	2,610	3 No	Football	PC	8%	3,800	No Ship	Mall	25	Operations	No	16	\$264,107	
SouthEast	1060	87,932	680	993	3 No	Soccer	PC	22%	8,885	No Ship	Stand Alone	29	Operations	Yes	8	\$100,466	
SouthEast	1061	27,742	440	720	3 No	Tennis	PC	53%	7,263	No Ship	Stand Alone	27	Operations	Yes	9	\$72,839	
SouthEast	1062	45,133	475	456	2 No	Tennis	PC	46%	2,276	No Ship	Stand Alone	22	Operations	No	7	\$46,147	
SouthEast	1064	76,267	547	2,460	1 No	Tennis	Mobile	42%	6,962	Ship	Mall	18	Marketing	No	18	\$248,927	
SouthEast	1065	86,726	625	1,966	3 No	WorkOut	Mobile	52%	11,552	Ship	Mall	28	Marketing	No	20	\$198,909	
SouthEast	1066	27,906	373	1,685	4 No	WorkOut	Mobile	33%	5,449	No Ship	Mall	26	Operations	Yes	29	\$170,547	
SouthEast	1067	58,056	368	760	4 No	WorkOut	Mobile	10%	7,016	No Ship	Stand Alone	21	Operations	Yes	22	\$76,861	
SouthEast	1068	25,952	380	1,189	1 No	WorkOut	Mobile	34%	2,956	No Ship	Stand Alone	29	Operations	Yes	8	\$120,327	
SouthEast	1134	51,299	710	1,224	4 No	Baseball	Mobile	31%	3,834	Ship	Mall	20	Operations	No	18	\$123,818	
SouthEast	1136	74,813	364	725	2 No	WorkOut	Mobile	2%	9,544	Ship	Mall	26	Operations	Yes	26	\$73,345	
SouthEast	1137	40,621	650	2,108	5 No	WorkOut	PC	31%	3,739	Ship	Mall	26	Operations	Yes	12	\$213,330	
SouthEast	1138	93,210	393	414	3 Yes	WorkOut	PC	18%	9,527	No Ship	Mall	24	Operations	Yes	20	\$41,897	
SouthEast	1139	17,818	423	1,835	4 Yes	WorkOut	PC	51%	3,131	No Ship	Mall	21	Customer S	Yes	26	\$185,727	
SouthEast	1141	9,545	520	1,979	1 No	Baseball	PC	50%	2,827	No Ship	Mall	24	Customer S	Yes	20	\$200,250	
SouthEast	1142	89,746	709	913	2 No	Baseball	PC	48%	2,987	No Ship	Mall	43	Customer S	Yes	17	\$92,421	
SouthEast	1143	5,368	506	1,894	3 No	Baseball	Mobile	6%	9,725	No Ship	Mall	61	Customer S	Yes	10	\$191,622	
SouthEast	1144	76,103	498	2,554	2 No	Running	Mobile	11%	6,169	No Ship	Mall	22	Customer S	Yes	30	\$258,440	
SouthEast	1145	20,417	533	561	5 Yes	Baseball	Mobile	28%	12,162	Ship	Mall	26	Customer S	Yes	12	\$56,723	
SouthEast	1146	53,031	400	1,520	5 Yes	Running	Mobile	27%	6,307	Ship	Mall	24	Customer S	Yes	27	\$153,849	
SouthEast	1147	83,985	650	811	4 Yes	Running	Mobile	15%	7,723	Ship	Stand Alone	26	Customer S	Yes	29	\$42,800,900	
SouthEast	1149	3,861	576	867	3 No	Baseball	Mobile	3%	4,574	Ship	Mall	25	Customer S	Yes	10	\$87,766	
SouthEast	1150	21,239	658	990	5 No	Baseball	Mobile	14%	3,229	Ship	Mall	30	Customer S	Yes	33	\$100,213	
SouthEast	1151	56,580	384	2,400	5 No	Running	Mobile	18%	8,538	Ship	Mall	26	Customer S	Yes	7	\$242,829	
SouthEast	1153	39,070	378	836	1 Yes	Running	Mobile	44%	5,313	Ship	Mall	27	Operations	Yes	31	\$84,578	
SouthEast	1154	49,999	469	351	3 No	Running	Mobile	65%	10,262	Ship	Mall	29	Operations	Yes	24	\$35,547	
SouthEast	1156	60,310	479	1,704	5 No	Running	Mobile	38%	8,928	Ship	Mall	29	Operations	No	11	\$172,445	

This is the final table which can be used for Data Analysis and the Data Warehouse manager can use this data which is clean and has no mistakes.

2) REGRESSION ANALYSIS

2.1) USING EXCEL-

STEP 1: From Data Analysis toolbar choose REGRESSION.

In the Input Y range, we select the column that we have to predict i.e Target Variable (MONTHLY REVENUE). In the X range we select all the columns that help in prediction i.e Driver Variables (App, Take Out Service, Area Income, Town Population, Outdoor Dining, Distance from Town Center) and click OK.

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.875598228					
R Square	0.766672257					
Adjusted R Square	0.751455231					
Standard Error	85521.9652					
Observations	99					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	6	2.21099E+12	3.68498E+11	50.38252693	5.67029E-27	
Residual	92	6.72889E+11	7314006531			
Total	98	2.88388E+12				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	288567.3571	50019.31254	5.769118814	1.06166E-07	189224.6848	387910.0293
App	-76373.7918	19164.45076	-3.985180309	0.000134882	-114436.0452	-38311.5384
Take Out Service	38400.52946	18111.07702	2.120278624	0.036673703	2430.367217	74370.69171
Area Income	0.547800806	0.399199191	1.372249292	0.173322398	-0.245043244	1.340644857
Town Population	0.833160385	0.140373235	5.935322234	5.12627E-08	0.554367023	1.111953747
Outdoor Dining	100934.4921	28050.49445	3.598314184	0.000518153	45223.78885	156645.1953
Distance from Town Center	-39449.73718	5999.968557	-6.574990653	2.89717E-09	-51366.19263	-27533.2817

STEP 2: Using the Formula, we calculate the Expected Revenue for the given values.

Formula = Coeff (Intercept) + App * Coeff (App) + Take Out Service * Coeff (Take Out Service) + Area Income * Coeff (Area Income) + Town Population * Coeff (Town Population) + Outdoor Dining * Coeff (Outdoor Dining) + Distance from Town Center * Coeff (Distance from Town Center).

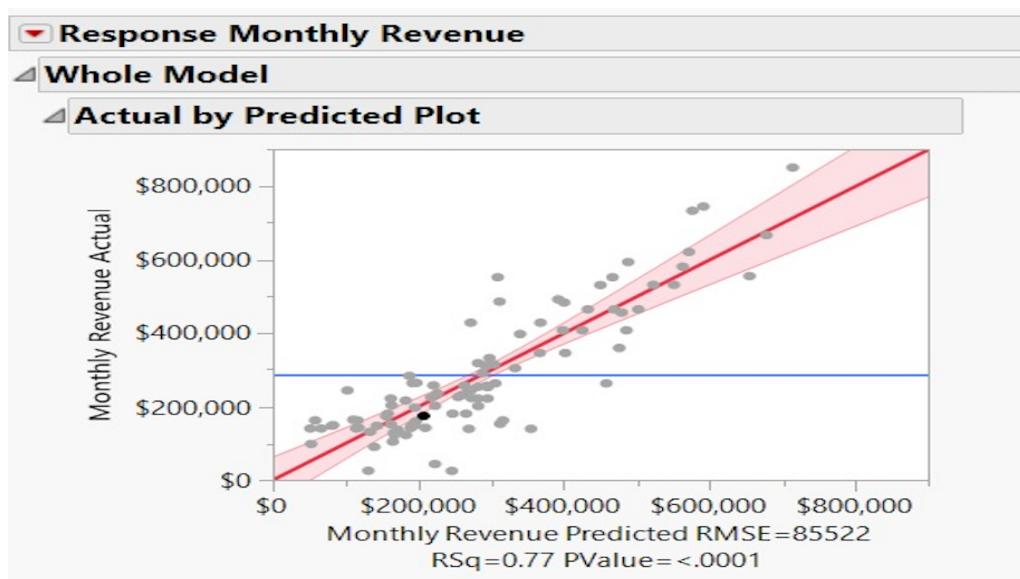
FORMULA = COEFF(INTERCEPT + APP*COEFF(APP) + TAKEOUTSERVICE*COEFF(TAKEOUTSERVICE) +.....DISTANCEFROMTOWNCENTRE*COEFF(DISTANCEFROMTOWNCENTRE)
--

Final Output							Distance from Town Center	Revenue
Restaurants	App	Take Out Service	Area Income	Town Pop	OutDoor Dining			
PPI	0	1	\$56,000	29,000	1	9	127693.24	
TRE	0	0	\$110,000	115,000	0	1	405189.15	
GGT	1	1	\$67,000	252,000	0	3	378903.95	
MND	0	0	\$105,000	71,000	0	2	326341.35	
WRT	0	1	\$67,000	150,000	1	1	550129.35	
GFR	0	1	\$74,000	120,000	0	7	191336.23	
WWW	1	1	\$130,000	75,000	1	2	406330.25	
QWE	0	0	\$51,000	100,000	0	4	229792.87	
FGR	1	0	\$45,000	47,600	0	3	158153.82	
SSC	0	1	\$72,000	91,000	1	11	109214.52	
SAE	1	1	\$85,600	68,000	1	4	297276.29	

2.2) USING SAS AND JMP –

STEP 1 – IMPORT DATA IN JMP.

Once we import the data from Excel it can be analyzed by the Fit Model tool in JMP. To start the analysis the output variables must be defined. The Response of Monthly Revenue is being measured against the App, Take Out Service, Area Income, Town Population, Outdoor Dining, Distance from Town Center.



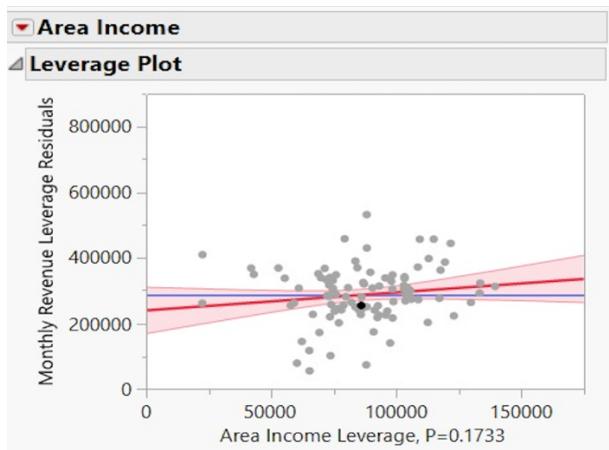
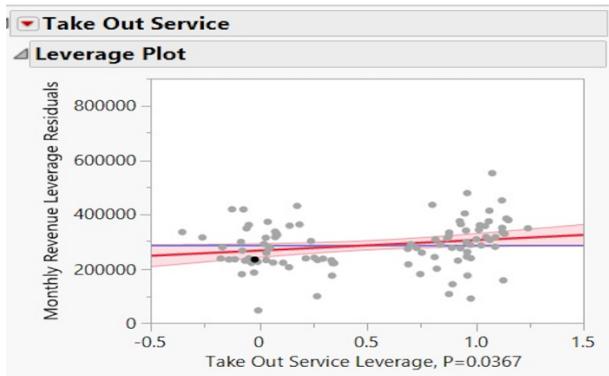
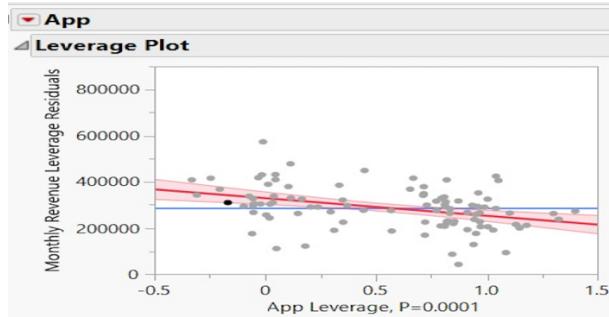
STEP 2 – ANALYZING THE DATA.

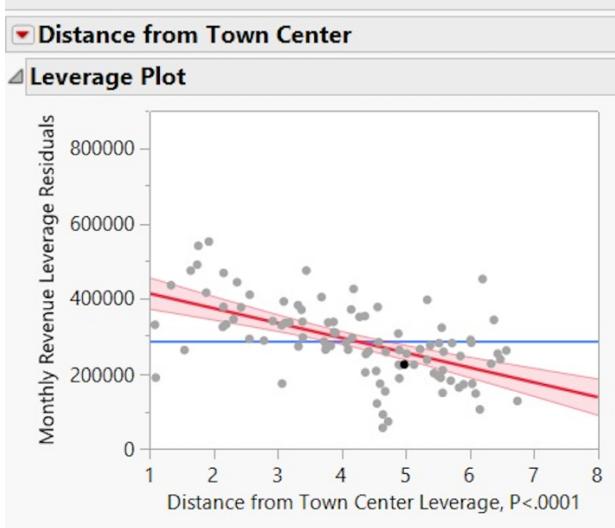
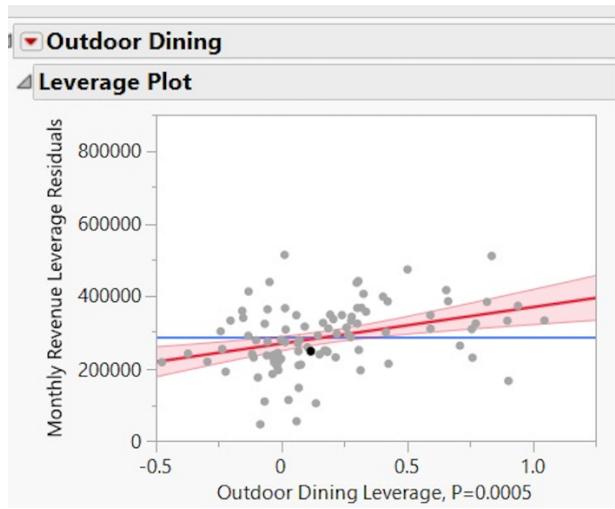
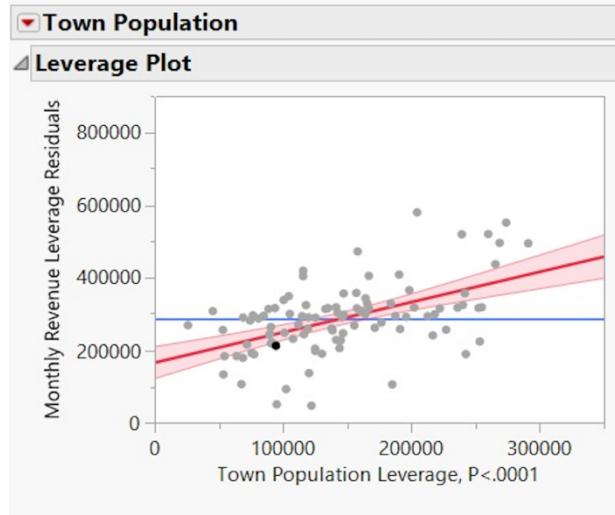
The total variation in our response values can be broken down into two components: the variation in the model and the unexplained variation or noise. From the chart above, the Monthly Revenue - Whole Model Actual by Predicted Plot output has a positive correlation and RSq=.77. The closer that RSq value is closer to 1 the data

falls almost on the straight line. Another measure of accuracy is the P value of a model. The lower the p value (below $<.05$ - .01) suggests that there is a higher effect on the data/population.

STEP 3 – ANALYZING THE RESIDUALS.

The lower the P value suggests that the null hypothesis is wrong, and that there really is an effect out there in the population.





STEP 4 – SUMMARY STATISTICS AND OUTPUT.

We can observe that the values are exactly the same that we observed while performing Regression in excel and we use the same formula to calculate expected revenue for the given restaurant data.

Summary of Fit					
	Source	DF	Sum of Squares	Mean Square	F Ratio
	RSquare		0.766672		
	RSquare Adj		0.751455		
	Root Mean Square Error		85521.97		
	Mean of Response		287303.2		
	Observations (or Sum Wgts)		99		
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	6	2.211e+12	3.685e+11	50.3825	
Error	92	6.7289e+11	7.314e+9		
C. Total	98	2.8839e+12			<.0001*
Parameter Estimates					
Term		Estimate	Std Error	t Ratio	Prob> t
Intercept		288566.89	50019.32	5.77	<.0001*
App		-76373.79	19164.45	-3.99	0.0001*
Take Out Service		38400.478	18111.08	2.12	0.0367*
Area Income		0.5478027	0.399199	1.37	0.1733
Town Population		0.8331605	0.140373	5.94	<.0001*
Outdoor Dining		100934.46	28050.5	3.60	0.0005*
Distance from Town Center		-39449.7	5999.969	-6.57	<.0001*

FINAL OUTPUT							
Resturants	App	Take Out Service	Area Income	Town Pop	OutDoor Dining	Distance from Town Center	Revenue
PPI	0	1	\$56,000	29,000	1	9	127693.2
TRE	0	0	\$110,000	115,000	0	1	405189.2
GGT	1	1	\$67,000	252,000	0	3	378904
MND	0	0	\$105,000	71,000	0	2	326341.4
WRT	0	1	\$67,000	150,000	1	1	550129.4
GFR	0	1	\$74,000	120,000	0	7	191336.2
WWW	1	1	\$130,000	75,000	1	2	406330.3
QWE	0	0	\$51,000	100,000	0	4	229792.9
FGR	1	0	\$45,000	47,600	0	3	158153.8
SSC	0	1	\$72,000	91,000	1	11	109214.5
SAE	1	1	\$85,600	68,000	1	4	2972776

2.4) NEURAL NETWORK –

Step 1: Import data from the excel file to JMP.

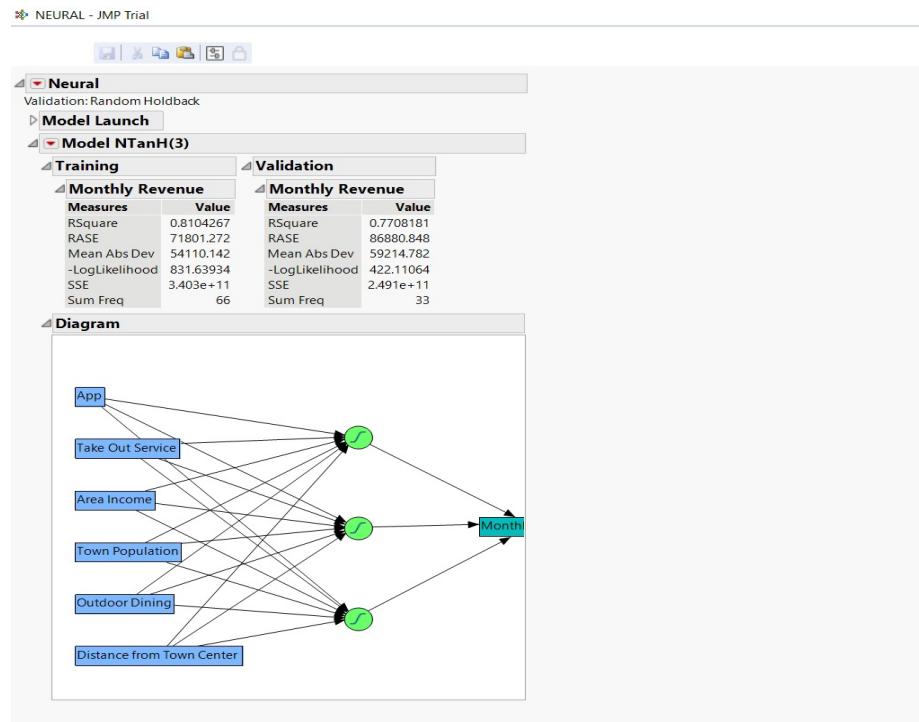
Step 2: Select Analyze, Predictive Modeling, Neural.

Step 3: Select Monthly Revenue as Y, Response.

Step 4: Select each of the factors as X, App, Take Out Service, Area Income, Town Population, Outdoor Dining, Distance from Town Center.

Step 5: Click OK and then Go. We used 3 hidden nodes for this example.

Step 6: To view diagram, click the red triangle next to Model NTanH(3), and then click Diagram.



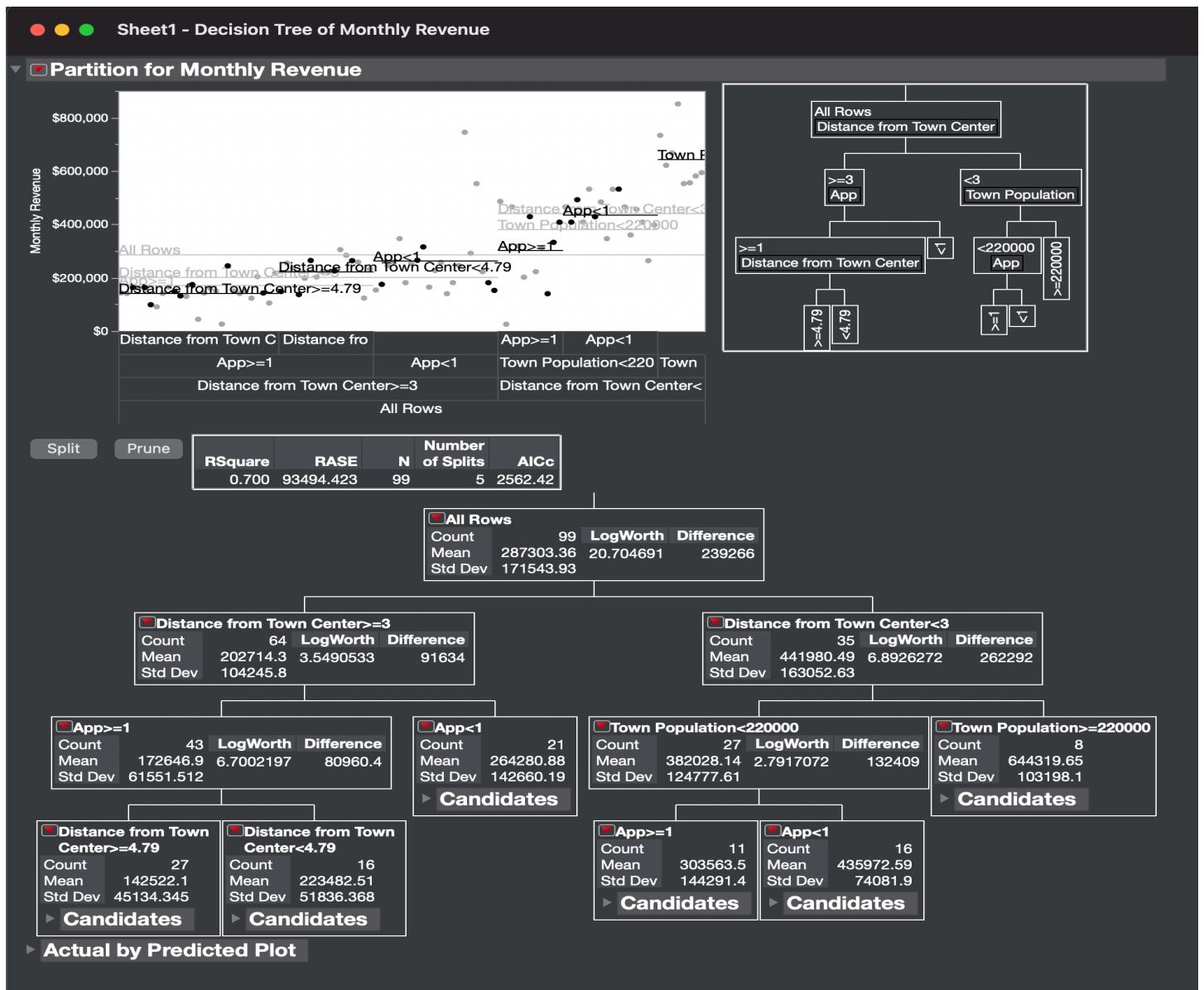
Step 7: The Prediction Profiler can be viewed using the same method as the diagram. This is interactive and can be used to show how each factor affects the Monthly Revenue price.



Our Regression Analysis has an RSquare value of 0.77, and our Neural Network has a similar RSquare value of 0.81. An RSquare value of 1.0 would indicate that 100% of the variation in the dependent variable is caused by the independent variables. Our RSquare values indicate that well over half (77% and 81%) of the variation in claims prices is caused by the seven factors we looked at.

When contrasting the outcomes, both the anticipated claims done through Regression Analysis and Neural Net are inside a similar reach with a minor distinction. The justification for this is on the grounds that the R^2 are inside a similar reach. When R^2 arrives at more like 1, the higher the rate that the driver factors are the immediate impact to the Target Variable.

3) DECISION TREE



Town Population<220000			
Count	27	LogWorth	Difference
Mean	382028.14	2.7917072	132409
Std Dev	124777.61		

PART 1 -

Using JMP, we created a segmentation model which partitioned drivers into groups based on certain characteristics. Since in the question it is mentioned that the town population is 21000 and the restaurant is at a location less than 3 miles from the center of a NJ town, hence we used these variables in our decision tree above.

RESULTS –

PART 2 –

<input checked="" type="checkbox"/> App >= 1	<input checked="" type="checkbox"/> App < 1
Count 11	Count 16
Mean 303563.5	Mean 435972.59
Std Dev 144291.4	Std Dev 74081.9

As we can see the Revenue when the App is being used is **\$303563.5** whereas when the App is not being used is **\$435972.59**. Now, if we add **100,000** that the tech giant is willing to pay to use the app the $(\$303563.5 + \$100,000)$ gives \$403563.5. Still not using the app Revenue is still greater by

$$(\$435972.59 - \$403563.5) = \$32,409.09$$

Therefore, it is not worth using the app.

